Handling Large-Scale Unit Level Data Using STATA Professor. Pratap C. Mohanty Department of Humanities and Social Sciences Indian Institute of Technology, Roorkee Lecture No. 22 Extraction in Stata-II

Welcome friends once again to our NPTEL MOOC module on Handling Large-Scale Data Using Stata. In the last lecture particularly we discussed extraction of the raw data ASCII format data to usable format in Stata. Broadly we try to discuss 3 types of data availability in social science and even in other field. They are either available in free format data, which we have already discussed, delimited format data and the third one is called fixed format data.

So, last two we discussed already. The free format data has the space delimiter. Space is used for delimiting the data or differentiating the variable from one to another. Then, the delimiter data sometimes available with tab-delimited data or with comma-delimited data. There are two ways. One is called, the data availability is in the form of dot csv or dot tab.

So, we are going to discuss the third format which is mostly used in our research and that is called fixed format data. The fixed format files are text files. But unlike CSV files there is no separator between variables. So, those seems very clumsy and looped together and but very interesting to observe and separate this data through another supplementary file we are going to discuss this.

(Refer Slide Time: 02:34)



Survey data often come in fixed format with one or more records per case and each variable in a fixed position in each record. So, that position is very clearly mentioned. For each variable fixed position space is clearly mentioned in the supplementary file we are going to discuss. This fixed format recognizes data items by column positions. The column position by their ranges also. The column positions are fixed for each row. We have row and column. So, the column position of that particular row is fixed in each particular row.

The data cannot be read directly, but a layout file is, as I just said couple of minutes back that it requires a layout file or a supplementary file, which explains how the data stored and most importantly how their position is defined.

(Refer Slide Time: 03:41)



And the advantage of fixed format files is that they are smaller by space, no space is wasted by the separator or even by the comma which you used in the earlier formats, but here no space is consumed because of that. The disadvantage is that they are completely useless without the corresponding data dictionary. So, data dictionary we will also discuss. If there are many variables to be read, it is efficient to write a separator file to provide necessary information and this file is none other than called data dictionary which I will be going to discuss.

Like data dictionary may not be useful if you have a very small set of variables. But if you have large number of variables usually it is like NFHS covers more than 5,000 variables. So, 5,000 variables are very difficult to enter manually. You need to create a manual dictionary first, then that dictionary is going to be useful every time. In every short span of time, you can extract the data and get your right formatted, extracted data for your use.

So, a data dictionary is defined as the file that explain how the variables are read and write. In India, most of the large scale unit level raw datasets are published in this particular format. Some examples are NSS datasets, economic census datasets etc. we are going to show it.

(Refer Slide Time: 05:23)



Let us understand the extraction of unit level data with a particular note that for this purpose we are using sixth economic census data that is the latest dataset on economic census is of sixth one, sixth round that is of 2013 and 14. The economic census provides the count of establishments including those in unorganized sector which have a huge contribution in Indian economy in terms of employment generations. It encompasses entire enterprises, economic census, entire enterprises in India.

(Refer Slide Time: 05:54)





6 0

🧼 🔟 👩

- 1 0 4 10 1 520 PM



So, let us understand how it is positioned in the original data and how it is available in the dataset from the ministry if you are interested in downloading. I will show you how it is available. This is the one I wanted to show you. Reports, then state wise data in ASCII format, we will also show you supplementary documents, these three are going to be very very useful.

Reports are important because you need to cross check, reports are derived by their own experts who has developed the data. They took the help of the statisticians and also team of economist for the economic related data. And accordingly they derived the national reports or state wise reports as well and those are usually the base results for us from that particular dataset.

Once you have extracted your data, you need to cross check whether you have correctly extracted or not, you need to cross check with the original reports. So, reports contain, the dataset, the reports of that is here. Let me just show you. It is opened. This is of 390 pages and all India report of the sixth economic census is. It gives many information with important results. Representative information some very stylized results are derived for us to compare. So, once you go through you can understand what this contains.

Similarly, I am going to show you all those detail later, but if you just have a look, let it be for Jammu and Kashmir, you need to open through, right click on it and open it, properties, open with is not there, then it is, I will show you in other file. It is there. So, this one is also, right click on it, select this, we are selecting this first, ok, then notepad. So, this is the data for you and we are going to use it. And this is the original data available by the ministry for the public use for the researchers with certain minimum agreements they do without any charge and these are freely available. But it is very difficult to read from this ASCII or this txt data and we need to extract all that information to a Stata format.

Mark carefully that there is no space in between, no comma in between, no extra space is consumed as per our earlier two formats. We are going to extract it based on certain instruction, I wanted to show you. Likewise, we have shown you Uttarakhand with the state code 05 raw data. Other state you can also get it on your own from the ministry website. So, you will get all the information in detail for extraction.

The supplementary documents are very very important. Within that layout file, we are going to guide you in a short while. This is going to be very very useful. It contains in excel format as well as in PDF format. This is what I am going to guide you in my couple of slides. So, let us have patience and I will guide you from the slide first and I will tell you how to read in between the lines. So, this is the one which is displayed for you, is important and that I have already shown to you.

The report file gives information about the data that includes sampling procedure, the main respondents, the key finding from the survey etcetera. The file also helps in matching the results of analysis done by the ministry and your results.

(Refer Slide Time: 11:01)

economic census data ha	ave sepa	rate d	data file	for each	n state.
News	Date modified	Los	Cree .		
		1700	~		
EC6A_STD1_JAMMU_AND_KASHMIR	07-06-2016 04:57 PM	DAT File	28,431 KB		
EC6A_ST02_HIMACHAL_PRADESH	07-06-2016 04:58 PM	DATHIE	23,350 KB		
ECEA_STUS_PURUAB	07-06-2016 04:59 PM	DATEL	85,712 KB		
	07-06-2016 05:01 PM	DATEIL	4,754 KD		
	22.06.2016 12.42 PM	DATEIA	65 075 VR		
FC6A_ST07_DELHI	07-06-2016 05:06 PM	DATFile	49 578 KB		
EC6A STOR RAJASTHAN	07-06-2016 05:07 PM	DAT File	1.63.982 KB		
EC6A ST09 UTTAR PRADESH	07-06-2016 05:09 PM	DAT File	3 78 581 KB		
EC6A STIO BIHAR	07-06-2016 05:11 PM	DAT File	95.709 KB		
EC6A_ST11_SIKKIM	07-06-2016 05:12 PM	DAT File	2,109 KB		
EC64_ST12_ARUNACHAL_PRADESH	07-06-2016 05:14 PM	DAT File	2,063 KB		
EC6a_ST13_NAGALAND	07-06-2016 05:15 PM	DAT File	3,452 KB		
EC6A_ST14_MANIPUR	07-06-2016 05:17 PM	DAT File	13,019 KB		
EC64_ST15_MIZORAM	07-06-2016 05:18 PM	DAT File	3,257 KB		
EC6A_ST16_TRIPURA	07-06-2016 05:21 PM	DAT File	13,411 KB		
EC64_ST17_MEGHALAYA	07-06-2016 05:22 PM	DAT File	5,979 KB		
EC6A_ST18_Assam	07-06-2016 05:24 PM	DAT File	1,14,983 KB		
EC6A_ST19_WEST_BENGAL	07-06-2016 05:26 PM	DAT File	3, 34, 500 KB		
EC6A_ST20_JHARKHAND	07-06-2016 05:29 PM	DAT File	36,178 KB		
EC6A_ST21_ODISHA	07-06-2016 04:21 PM	DAT File	1,18,317 KB		
EC64_ST22_CHATTISHGARH	08-06-2016 12:37 PM	DAT File	43,821 KB		
EC6A_ST23_MADHYA_PRADESH	08-06-2016 12:40 PM	DAT File	1,21,928 KB		
EC6A_ST24_GUIRAT	08-06-2016 12:44 PM	DAT File	2,25,030 KB		

This is the one I have already shown you. These are in ASCII format. We need to extract it for our use. The sample we have shown you.

(Refer Slide Time: 11:09)



Each row in the dataset represent a single observation with varying information about its unit that each row here we will have to write each row is representing information about a single individual, single respondent. But here in our sixth economic census data, it is not individual, it is the enterprise. If I do any analysis, I will talk about the enterprise information. And without extraction it is impossible to do any meaningful analysis of this dataset.

Name	Date modified	Туре	Size
Ec6A_Report_Tabulation_Plan	17-08-2020 04:56 AM	File folder	
EC6A_Schedule	17-09-2019 04:20 PM	File folder	
EC6A_UNIT_LEVEL_Data_Layout	17-08-2020 04:57 AM	File folder	
District_Code_List	16-06-2016 04:52 PM	Microsoft Excel 97	149 KB
🔂 Guide_for_enumerators_and_supervisors	06-06-2016 06:40 PM	Adobe Acrobat D	902 KB
Instruction_Manual_EC6_2012_2013	06-06-2016 06:39 PM	Adobe Acrobat D	1,162 KB
🚯 Master_Directory_6EC	17-08-2020 05:15 AM	Microsoft Access	9,38,720 KB
NIC2008_3_DIGIT_CODE_LIST	20-06-2016 05:30 PM	Microsoft Excel 97	48 KB
🗑 State Code for 6th EC	20-08-2015 03:12 PM	Microsoft Word 9	47 KB

(Refer Slide Time: 11:53)

So, this is what I have already shown to you. In addition to that, the three files, important files I have shown you. You might be confused, which code is it like 05, though we know that it is for Uttarakhand, you might be confused like district code list is given to you. Once you click it, the code number is visible to you. Similarly, NIC classification, enterprises without NIC classification is meaningfulness. There must be some standardization, some classification required. Three-digit classification information is given in sixth economic census dataset.

(Refer Slide Time: 12:34)

- □ The most important file that is key to your work for extracting data are provided in the folder "EC6A_UNIT_LEVEL_data_Layout" that contains layout file in excel and pdf versions.
- □ The layout file guides how the data is arranged, which column gives what information.





And similarly other information, instruction manual is there if you read you will be guided accordingly. So, the focus is to understand the layout file which I just shown you couple of minutes back. And the most important file that is the key to your work for extracting data are provided in the folder called EC6A which I have just shown you, it is here. This is the one EC6, economic census 6A underscore unit level data layout that is the one is going to be very useful that contains layout file in excel as well as PDF version.

I have already shown you, this contains both the information. The layout file guides how the data is arranged and which column gives what information.

(Refer Slide Time: 13:25)

Variable	able Column No. as in Schedule		Column Column No No. as in Position in Byte Schedule Electronic Data file		Remarks	
		Starting	End			
State / Ut		1	2	2		
District		3	4	2		
Tahsil/Taluka/P.S./Dev. Block Circle/Mandal		5	7	3		
Town/Village		8	11	4		
Ward Code No. (Only for Town)		12	15	4		
Enumeration Block No.		16	19	4		
Extended Enumeration Block No.(EBX)		20	21	2		
Use of Census House/Structure Code	2	22	22	1		
Inside HH	8	23	24	2		
Broad Activity Code	12	25	26	2		
NIC 3 digit Code	13	27	29	3		-
is it a handloom/handicraft activity?	14	30	30	1		- 11
Contraction Code	15	31	31	1		
Social Group	10	32	32	1		
Religion	18	34	34	1		
Nature of Operation	19	35	35	1		
Major Source of Finanace	20	36	36	1		
Male Hired	21	37	41	5		
Female Hired	22	42	45	4		
Adult Alexa blood	22	46	47	2		
Male Not-hired	23	40				
Female Not-hired	23	48	49	2		
Female Not-hired Total (hired + not hired)	24	48	49	2		3
Nake Korhired Female Rothired Total (hired + not hired) Sector	23 24 25	48 50 56	49 55 56	2 6 1		
Water Morning Female Not hired Total (hired + not hired) Sector Women Harp 1 100 100 100 Record Layout of Sixth Economic Cens Variable	us Schedu Colum No. as Schedu	ule 6A b in in Pule Ile E	49 55 56 ased Colur ositio	2 6 1 Unit L nn n in pnic	evel Re No of Bytes	cords Remarks
Water Morning Female Not hired Total (hired + not hired) Sector Women Harp 1 100 100 100 Record Layout of Sixth Economic Cens Variable	23 24 25 Us Schedu Colum No. as Schedu	48 48 50 56 56 10 10 11 11 11 11 11 11 11 11 11 11 11	49 55 56 ased Colur ositio lectro Data	2 6 1 Unit L nn n in pnic file End	evel Re No of Bytes	cords Remarks
Water Monthined Female Not hired Total (hired + not hired) Sector Sector Record Layout of Sixth Economic Cens Variable State / Ut	23 24 25 us Schedu Colum No. as Schedu	48 48 50 56 10 10 10 10 10 10 10 10 10 10 10 10 10	49 55 56 ased Colur ositio lectro Data 1	2 6 1 Unit L nn n in nic file End 2	evel Re No of Bytes	cords Remarks
Make Not-Interd Fenale Not-Interd Total (hired + not hired) Sector Record Layout of Sixth Economic Cens Variable State / Ut District	us Schedu Colum No. as Schedu	ule 6A b so so so so so so so so so so so so so	49 55 56 ased Colur ositio lectro Data 1 1 3	2 6 1 Unit L nn n in nic file End 2 4	evel Re No of Bytes	cords Remarks
Male Mothined Fenale Not-hired Total (hired + not hired) Sector Fenale Not hired Fenale Not hired	us Schedu Colum No. as Schedu	ule 6A b so so so so so so so so so so so so so	49 55 56 ased Colur ositio lectro Data trting 1 3	2 6 1 Unit L nn n in pnic file End 2 4 7	evel Re No of Bytes 2 2 2 3	cords Remarks
Midde Nort-Nired Fenale Nort-Nired Total (hired + not hired) Sector	us Schedu Colum No. as Schedu	ule 6A b so so so so so so so so so so so so so	49 55 56 S6 Colur ositio lectro Data tring 1 3 5 8	Unit L Unit L End 2 4 7	evel Re No of Bytes 2 2 3 4	cords Remarks
Mode Vorhied Fernale Not hired Total (hired + not hired) Sector Record Layout of Sixth Economic Cens Variable State / Ut District Tahsil/Taluka/P.S./Dev. Block Circle/Mandal Town/Village	23 24 25 us Schedu Colum No. as Schedu	ule 6A b so so so so so so so so so so so so so	49 55 56 ased Colur Data 1 3 5 8	2 6 1 Unit L nn n n nin c file 2 4 7 11	evel Re Bytes 2 2 3 4	cords

I have already opened. This is the sample of that. That I already opened to you. Let me guide it from here or from that once again for you. You can open in excel or in PDF. It does not matter. How to go for it? look at the state or union territory. The variable is occupying space starting with the first column is occupied for this till the second. The total digit number of bytes it consumed is 2. Till the 2 space in the entry of our raw data, in the raw data there are so many, without any space without any column, which I have already shown to you.

16 19

20 21

25 26

27 29

22 22 1

23 24 2

30 30 1

31 31 1

2

8

12

13

14

15

4

2

2

3

- 📱 📀 4 🗊 h 👷 526 PM

Enumeration Block No.

Broad Activity Code

NIC 3 digit Code

Ownership Code

🧈 🗋 🌀 🔛

Inside HH

Extended Enumeration Block No.(EBX)

Use of Census House/Structure Code

Is it a handloom/handicraft activity?

The first two for your sample information once again, I know that you might be confused for it, it is here. So, I will show you again, like this is the information we have already shown to you.

Now here our next one is like, here the first two look at this, for us, 05 we have already filtered. So, I am showing the 05 here. First two represents the state and union territories. The sample we are showing it for a particular state. So, if you see the original data and the file we are now extracting is the state that is Uttarakhand sub-file. Similarly, for Delhi you will get another code, for Orissa you will get another code, for Bihar you will get, the starting two code is for that only.

Then what comes next. The third position and the fourth position occupied for district and it occupies 2-byte space. Similarly, you can read in between the lines town, village, from 8 position to 11 position, if we can, in our command if you can attach this particular position then it will convert our data correctly. And this is very very essential. The byte position should be correctly written. If by any chance you have done a mistake here, at the time or result it is very difficult to track.

Though it might be extracted with certain other position, but other position will be named with another variable name and it might be problematic or it may be the case that your extraction is not successfully completed. However, we always suggest that you must enter these carefully while extracting. Then what comes next for us is the following. We will explain one by one and how to extract it.

(Refer Slide Time: 16:15)





So, this we have already guided. Another tip is here for you. Always read the layout file very thoroughly as I mentioned. Another important file is schedule file. With schedule we are going to extract it. It describes the sequence of questions upon which information is collected. It also helps in labeling the variable that the variables values. So, schedule file, generally we called the questionnaire, the questionnaire gives all the set of information. We have already shown you earlier. But if you are again confused it is there, the schedule file.

Here is the schedule. The schedule files guide you, systematic direction for the questionnaire. It is being opened. Then instruction manual it might be also at the end of that as well. It has some content. Probably it is taking little time to open. Do not get confused by this. Since it is not opening, it might be taking little more time. But for you that it is the questionnaire, you need to read very thoroughly that what question follows next. Accordingly, the variables are also entered with their space. Understanding question from the questionnaire is very very important.

(Refer Slide Time: 17:512)



Let us understand creating a dictionary file which is essential in case of this fixed format without that also you can operate. We are going to guide you. Let us launch the Stata and open a do-file editor from there we can do accordingly and third one write down the data structure as specified in the layout file. Here we have mentioned for you. You just mark carefully in fixed dictionary then all those entries. look at here this data since is in string format and we have written str then state with their position 1 to 2 then district then 3 to 4 then 5 to 6.

If you want small set of variables, you need not mention everything. You simply mention those variables for your use for extraction. No need to include everything here. At the end, you have to, in the dictionary dot dct file or dictionary file you need to end with this bracket. Then only it will ready. We are using string type of consistent extraction. You can use other preferable storage type as well. Let me show you that, the dct file which we wanted to show you here from the Stata, you can open the do-file you can create here and accordingly you can go. I can open that, you can type accordingly and save as in a dct, but I am just opening it. File open simply.

But dct has to be there. All file let me open that, it will open automatically. So this is the format for the dct. In fixed dictionary you need to write down and then starting with a bracket and go by this ordering. Try to follow this ordering as per the standard suggestion. It will give the extraction. what I will do, I will guide it from the PPT.

(Refer Slide Time: 20:28)





Otherwise, dct file can be also opened, once you have opened the dct file or then the do-file once you have typed your command then you just save it. Then save as the way I just said here that you simply file save as. If it is not in dct file initially in its do command, you need to save it save as dct. You need to go for dct, like here dictionary. You have to save it. Then only it will read. I will guide you from the PPT. likewise, dct has to be entered. Then you save it.

(Refer Slide Time: 21:13)



There are 3 methods of extracting the fixed format data. First is, once your dct file is ready, you can take the use of dct file. You launch the Stata file then import then text data in fixed format, it

has to be mentioned. Then browse the particular dct file, I will tell you right now, if you have created a dct file. Else you simply click on the specifications, specification I will also guide you, type a chosen variable name along with storage type or the column position accordingly if you can define manually.

Browse your dataset file that is which dataset we are going to extract, isn't it, then file name, browse the file name of that particular dataset which you wanted to extract as we have not specified it in the dictionary file if it is not specified. If it is specified, then dictionary file will automatically read the file name. The check on replace data memory if there is any, already opened data file then you submit okay. I will show you right now.

(Refer Slide Time: 22:26)

Jata SE 151 In 1561 Data Graphics Statistics Liker Worden: Helm	
Coro. Chi-O with m A. O	
Save Chiefs and an and an and an	
Save as. Cb1-Shift+5	Vanables
. label define activity_group 1"manufacturing" 2"trade" 3"services"	Effer variables here
Da.	Name Label
Fireame	entid ENTID
Change working directory. activity grou	state State
p Freq. Percent Cun.	b2_q204
	67,6710
Import Excel spreadsheet ("xis" also) 20.51 20.51 20.51	02,0211 b2 o212
Export Text data (delimited, "csv,) 40.54 100.00	b2 o213
Print Text data in fixed format	b2 a216
Example datasets Text data in fixed format with a dictionary PD.00	b2,q218
Recent files Unformatted text data	b2_q219
Exit SAS XPORT	
Tabel value activity, group a	
insheet using G/WSS1-csv.c Haver Analytics database	
import delimited G\NSS1OOBC data source	Properties
import delimited G\NSSI4 dBase (*dbf) y, clear	8
import delimited G(NSS) (10 vars, 72,528 obs)	 Variables
import delimited G(NSS1	Name entid
no; data in memory would be lost	Label ENTID
r(4):	Type double
incode deligited delighted and delight	Format %10.0g
(10 vars, 72.528 obs)	 Value label
	Notes
. import delimited G:\NSS1-csv.CSV, clear	- Deta
(10 vars, 72,528 obs)	Label
	Notes
	Variables 10
	Observations 72,528
CONTRADO	Sue 138M
	Memory 64M
	Sorted by
aa15	CAP NUM
	535 Pb

T 8 × Tota	1 72,528 100	0			 Variables 	
nds here	fine activity group 17m	ufacturing" 2"trade" 3"services"			+ Filter varia	
.K					Name	Labe
ice_dataset_N ta activ	/ity_group				entid	ENTIC
e b2_q218 'co activity_c	iton				state b2 o.204	State
rple data_set.d.	p Freq. Per	ent Cun.			b2_q210	
manufactu:	ring 20,680 2	.51 28.51			b2_q211	
ti ti	rade 22,447 3	.95 59.46			b2_q212	
ctivity_group	29,101	infix - Import text data in fixed format			b2_q213	
ctivity_group	otal 72,528 10	.00 Main it/in			b2_0210 b2_0218	
activity.group	alue activity group activ	ty group B lise dictionary file			b2,q219	
roup		6	Browse.,			
ctivity_group a	using G:\NSS1-csv.csv.	ear Specifications				
G(NSSI-GVC. (10 vars,	72,528 obs)					
nited G/WSS14	alimited G-\NSS1-cev CS	Examples			Properties	
nited GNNSSI (10 vars,	72,528 obs)	rate 1-4 speed 6-7 acc 9-11			A Variables	
nited GI/NSS1		2 lines 1: id 1-6 str name 7-36 2:	age 1-2 sex 4		Name	
no: data i	in memory would be lost	Text dataset filename: (required if not	specified in dictionary)		Label	
r (4) ;			Browse		Type	do
imort of	felimited G:\NSS1-cev cs	clear			Format	
(10 vars,	72,528 obs)				Value label Notes	
(·	alinitad G-\NSS1.com of	elur 000	X Cancel Submit		4 Data	
(10 vars,	72,528 obs)				Filename	
					Label	
					Notes Vanables	
					Observation	
Command					Size	
					Memory	64
III. Gaphics Satisfics User W	ndow Help				- 2 0 (b h
D ≪ D () a Gapha Sastra Lier W 3 · ⊥ · ∬ · ⊠ ⊒ 0 0 T 4 × Tote	ndow Help	•			 Variables 	b N
C C	ndow Help 0 11 72,528 100 51a Activity group 170	0 0			Variables	to h
BL Gopha Sata Der W T - L - Z - 2 - 2 - 0 T + X - Tota shirer 0 . 1abel de	ndow Heip 0 11 72,528 100 fine activity_group 1'm	0 alastutiop D'tuder D'annicar			 Variables Variables Fitter varia Name 	to here toles here
ta Gapha Sontha Use W S - ta - t	ndow Help O NI 72,528 100 Afine activity_group 1*mw rity_group	9 alastuise Disate Dianesaar	_		 Variables Fitter varia Name entid 	to ha obles here Labe ENTIL
La Gapha Sonta User W S - M - Z - Z - Z - Toto no tere 0 . Ideal d . Idea	ndow Help 0 11 72,538 100 stine activity_group 1'mw Aity_group prov	0 ulatituisy: 2'titala: 2'arritas:	_		Variables Variables Fiter varia Name entid state	to here Labo ENTI State
Capita Satina Der W Capita Satina Der W T 2 X T 2	ndow Hep O ii 72,528 100 fine activity group 1'ms ntr _g group p freq. Per	0 alatharing: 2°teader 2°mercianer all Come	_		 Variables Filter varia Filter varia state state state 	Dies here Labi ENTI State
La Gopha Sonna Der W S - da - Z - da S - O V 2 X not tere V 2 X - Label d - ta activ - ta - ta	0000 HHD 0 1 72,528 100 0 11 72,528 100 0 11 72,528 100 0 11 72,528 100 0 11 72,528 100 0 11 72,528 100 11 72,528 100 12 72,528 100 10 72,528	e efectusion: 2"teader 3"services" e Com e Com e Com e compare + Remonitor Dia (5)		• 4 Sect Remode Dut	 Variables Fiter varia Name entid State 2, 204 b2, 201 b2, 201 	to Participation Provided Prov
Copies Sense Sense Copies Sense Sense Copies Sense Sense Copies Sense Copies Sense Copies Co	Non Non Non 0 1 72,528 100 0 1 72,528 100 0 1 72,528 100 0 1 72,528 100 0 1 72,528 100 0 1 72,528 100 0 1 72,528 100 0 1 72,528 100 0 1 72,528 100 0 1 72,528 100 0 1 72,528 100 0 1 72,528 100 0 1 72,528 100 0 1 72,528 100 0 1 72,528 100 0 1 72,528 100 0 1 72,528 100 0 1 72,528 100	0 ufactualing: 2"teads": 3"services" (Com (Com (Com) (Co		• + Senti Removati Qua	 Variables Variables	to Participation
A and a sense to	Image: Notice Project Image: Notice Image: Notice<	0 afattuilige "2"tesfer" 3"services" Com • Compare + Removable Dis (6) Organe - New Color • Services - New Color	> 	• + Secti Resource Data	 Variables Variables	to Participation
A Grants Statistics A Grants Statistics A Grants A Gr	Image: Non-Weight Control Image: Non-Weight Contro Image: Non-Weight Control Ima	e efectuciogi 2°tesder 3°services* © Con © Con © Con © Consoler - Remonite Dix (5) Operar - New Killer © Rent Plans - New New Coller	• Gire rooited	• • • • Secti Remote Due	 Variables Variables	to In Labo ENTI State
Control Contro Control Control Control Control Control Control Control Control Co	NON NED 0 0 1 72,338 100 0 0 0 1 72,338 100 1 72,338 100 1 72,038 100 1 72,038 100 1 72,038 100 1 72,038 100 1 20,400 1 1 22,401 1 1 22,401 1 1 22,401 1 1 12,538 1	0 vlaturing 'P'teade' P'services' Come - Compare - Removale Dia (6) Operate - New Using Spectr Rem ' target Spectr Rem ' target	• Our modified 71/2000 100 m or provided	• • • See See See See See See See See Se	A 2 0 4 A 2 0 A 2 0	It is here Labor ENTI State
A capital Sensitian Core W Core Core Core Core Core Core Core C	Image: Note of the section o	e efecturinge "2"tender "3"eerosices" Com Com Com Com Com Com Com Com	Cur mosted 7/12/2020 ID IA7 4/2020 ID IA7			It is in the second sec
Compare Source Core W Compare Sourc	Non Nep 0 0 1 22,528 500 1 22,528 500 1 22,528 500 1 22,528 500 1 22,528 500 1 22,528 500 1 22,407 10 1 22,407 10 1 22,407 10 1 22,407 10 1 10 10 1 10 10 1 10 10 1 10 10 1 10 10	e efectusion ² 2 ⁺ teste ⁺ 3 ⁺ services ⁺ Com Com Com Com Com Com Com Com	Der mobilet 711/2008 I 939 crazo 538 m Vacaujonan 1972/2009 I 18 m	• • • • See the function of the • • • • • • • • • • • • • • • • • • •	Variables	Diffs here Labo State
Company Section One W Company Secti	Image: Non-Section 2006 Percent PercentPercent Percent PercentPercent Percent	e afatturior 2*teat* 3*services* Com Com Com Com Com Com Com Com	Currented 7711/2008 10 50 400000 10 50 400000 10 50 400000 10 10 7171/2000 10 11 7171/2000 10 11 7171/2000 11	+ + Santh Remote Dat = ++ Santh Remote Dat = 	 Variables Variables	Difference in the second secon
A Granto Statistic Core W A	Image: Note Note 0 1 72,328 100 1 1 72,328 100 files activity_group Treq. Per 1 72,528 100 1 22,640 2 1 22,640 2 1 22,643 10 1 22,643 10 1 22,643 10 1 1,9,528 10 1 10,528 10 1 10,528 10	e efectualing: 2"teade" 3"services" e Com e Com e Com e Com e Com e Source - Remosite Dia (5) o gana - New Iode e Source - Remosite Dia (5) e Source - Remosite Dia (5) e Source - Remosite Dia (5) e Source - Remosite - Remosite - Remosite - Dia (5) e Source - Remosite - Remosite - Dia (5) e Source - Remosite - Remosite - Dia (5) e Source - Remosite - Remosite - Remosite - Dia (5) e Source - Remosite -	Oue modified ZU12000 IB IB P AU2000 IB IB P	for Secto Removale Dat Secto Removale Dat Sec Sec Secto Removale Dat Sector	✓ Variables ✓ Variab	Dies here Labies State
Control Southern Control Market Control Southern Control Market Control Southern Control Market	NON Nep 0 0 1 72,538 100 1 72,538 100 1 72,538 100 1 72,538 100 1 72,538 100 1 72,538 100 1 10,600 1 1 20,600 1 1 22,640 1 1 12,528 100 1 12,528 100 1 12,528 100 1 12,528 100 1 12,528 100	0 vlaturing" P'teste" P'services" Correct: Acrosofte P services Correct: Acrosofte P servi	Currential Currential Acacitomat	Figure Sector Removale Data Figure Sector Removale Data Figure Sector Removale Data Figure Sector Removale Data Figure Removale Data Figure Removale		D No.
Constant Section Over W	Image: Note of the sectivity group Image: Note of the sectivity group<		Our mosted 7/12/000 10 kP encode state control to the control to t	Par Seret Remonstrate Data Seret Remonstrate Data Ser Tope Ser Top Setter No Top Setter More Setter More Setter Ser Setter Ser Setter Set Setter	Variables	to N
Compare Source Our W Compare Source O	None Neg 0 0 1 32,538 500 1 32,538 500 1 32,538 500 1 32,538 500 1 22,640 1 1 22,640 1 1 22,640 1 1 22,548 10 1 22,548 10 1 10,521 10 1 10,521 10 1 10,521 10 1 10,521 10 1 10,521 10 1 10,521 10 1 10,522 00 1 10,522 00 1 10,522 00 1 10,522 00 1 10,524 00	e efectuciog" 2"tesér" 3"services" efectuciog" 2"tesér" 3"services" effectuciog" 3"services" effectucioge 3"services" effectucioge 4"services" effectucioge 4"services"	Gre mobiled Transition Transition Control (1997)	Iserth Remote Date Serth Remote Date Serth Remote Date Type Sore Type Sore The folder The folder The folder The folder Det Type Det Type	Variables	to N labe ENTI State
A serie of the series of the s	Image: Non-Section (Non-Section (N	O Afstücior 2*tede* 3*services* Open Com Com Com Com Com Com Com Co	Cue mostree 7/11/200 10.07 4/2000 10.07 4/2000 10.07 7/12/2000 10.17 7/12/2000 10.17 6/17/2000 10.11 A	+ s Santh Remote Dat - + s Santh Remote Dat - + s - + +	Vanables VVanables VVanabl	to N Labo ENTI State
A Granto Statistic Core W A Granto Statistic Core Batereage Attractoreage Attractoreage Attractoreage Attractoreage A	Image: Note of the second se		Cue modified 7/12/2020 Io Rav exceed you want to concern a second second second concern a second second second concern a second second second concern a second second second second second concern a second second second second second concern a second second second second second second concern a second second second second second second second concern a second second second second second second second second concern a second	Part Remote Control Contr	Variables	to IN Laber State
Compare Source Core W Compare Source Core Source	NON Nep 0 0 1 72,538 00 1 72,538 00 1 72,538 00 1 72,538 00 1 72,538 00 1 72,538 00 1 10,660 1 1 22,461 0 1 72,538 10 1 10,553 00 1 72,538 10 1 10,513 evr cer, er 1 72,528 00 1 10,513 evr cer, er 1 10,526 0.0013 1 10,526 0.0014 1 10,1266 0.0014	0 vistuaion - Pisade - Pisarrices - vistuaion - Pisade - Pisarrices - Operate - New lotter Sector Res - Marc - Compare - Comp	р — — — — — — — — — — — — — — — — — — —	In the second and the second an	Variables	to IN Laber State
A series of the series of	Image Image Image 0 0 0 1 72,538 100 0 0 0 11 72,538 100 0 0 0 10 72,938 100 11 72,938 10 11 22,401 10 12 22,401 10 13 72,538 10 141 72,538 11 152,528 10 12,528 152,528 10 12,528 152,1528 10 12,528 152,1528 10 12,528 152,1528 10 12,528 152,1528 10 12,528 152,1528 10 12,528 153,11446 10,1531-497 10 163,11449 10,1541 10,1531		Core modified Trailogo is the P Core modified Corection is the P Corec		Variables Variables Variables Variables Variables Variables Variables Variable Variab	to No Internet State State
Compare Source Our W Compare Source O	Image Image 1 32,538 500 1 32,538 500 1 32,538 500 1 32,538 500 1 22,640 10 1 22,640 10 1 22,640 10 1 22,640 10 1 22,640 10 1 22,528 10 1 10,521 00 1 1,522 00 1 1,522 00 1 1,522 00 1 1,522 00 1 1,524 00 1 1,524 00 1 1,524 00		•	fai Search Removative Duel for Search Removative Duel	Vanables Vanables Vanables Vanables Vanables Vanables Vanables Vanables Vanables Vanable Vana	to Ite
Control Contro Control Control Control Control Control Control Control Control Co	Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1		Currentee 7/11/2008 10:89 4/2000 20:89 7/21/2008 10:81 7/21/2009 10 7/21/2009 10 7/21/2009 10 7/21/2009 10 7/21/2009 10 7/21/2009 10 7/21/2009 10 7/21/2009 10		Vanables Vanabl	to Para
A Grants Control Core W A Grants	Image: Note Image: Note Image: Note 72,328 100 Image: Note 72,328 100 Image: Note 72,328 100 Image: Note 72,328 100 Image: Note 22,441 100 Image: Note 22,441 100 Image: Note 22,441 100 Image: Note 72,528 101 Image: Note 72,528 101 Image: Note Note 100 Image: Note		Cover most field 7/12/2020 10 km exceed you want to cover most you want to cover most wit 7/2/2020 12 21 A		Variables	to Participante de la construcción de la construcci
Compare Souther Core W Compare Souther Core Sout	NON N(D) 1 72,528 100 1 72,528 100 1 72,528 100 1 72,528 100 1 72,528 100 1 72,528 100 1 72,528 100 1 10,660 1 1 22,647 100 1 72,528 100 1 10,553 100 1 10,553 100 1 10,553 100 1 10,553 100 1 10,553 100 1 10,553 100 1 10,554 00 1 10,554 00 1 10,554 00 1 10,552 00 1 10,552 00 1 10,552 00 1 10,552 00 1 10,552 00	Construction of the second sec	Core modified Y110000 1949 Corector 194	Provide Control of Control o	Vanables	en la
Command	Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1 Image: Non-Section 1		Core modified Training of the second	++ Senth Remotely Call - ++ Senth Remotely Call - - + -	Variables	er Bites here Extension State
Command	NON Megn 1 22,528 000 1 22,528 000 1 22,528 000 1 22,640 101 1 22,640 101 1 22,640 101 1 22,640 101 1 22,528 101 1 22,528 101 1 22,528 101 1 10,528 001 1 10,528 001 1 10,528 001 1 10,528 001 1 10,528 001 1 10,528 001 1 10,528 001 1 10,528 001 1 10,528 001		Core modified 7712020 10 87 4712020 10 87 4712020 10 87 4712020 10 87 5004MMS 7712020 1271 A 4172020 1271 A	fai Sarch Remote Puel for Sarch Rem	Variables Variables Variables Variables Variables Variables Variables Variables Variable Varia	enters here inters

Impact Note Impact Note Impact Note </th
Antimit Image: Status and Status an
* C Sportsverskytter ** C Sportsverskytter ** Kittistig, green **
Mittal de l'autor d'autor d
* (Stepper Alexanda, NAMOR Alexanda, ganda, alexanda, ganda, g
NY MUM MY MUM and constant, grade and reg acting, grade and reg acting, grade and reg acting, grade and reg acting, grade and reg, grad
And grad And grad de de ga key grad de de de ga key grad de de de ga key grad de de de de de ga key grad de de de de de ga key grad de de d
Introduction of the second
ating mon ating mon
Method method water, group, and water, group, and water water water, group and grou
<pre>status</pre>
Anter und gold state Und gold Sta
our en
And Market Collins of Collins
verdemed GMS1: - uppet dislated 6 (MS1-exc CP), is used tail, in memory would be last (14); - uppet dislated 6 (MS1-exc CP), is used balance 3 and used in the MS1 and Msanath Mare, 1994; - uppet dislated 6 (MS1-exc CP), is used balance 3 and used in the MS1 and Msanath Mare, 1994; - uppet dislated 6 (MS1-exc CP), is used balance 3 and used in the MS1 and Msanath Mare, 1994; - uppet dislated 6 (MS1-exc CP), is used balance 3 and used in the MS1 and Msanath Msanath Mare, 1994; - uppet dislated 6 (MS1-exc CP), is used balance 3 and used balan
Provide dilution de (Millice Control of the Co
r (1) - pert dilatid () Wist-cer CP, 15 verz, 72,52 da) - pert dilatid () Wist-cer CP, - p
state de lastet de Justet
115 varse, 72,528 dba) - sport dilatifie & Wissi-cerv CEV, 10 varse, 72,528 dba) Command
Lippet deliaited 6.Wss1-eer.CP7, (1) vers, 72,518 das) Command
Command Comman
Command Com
Command 1 </td
Command Comman
S → D → D → D → D → D → D → D → D → D →
8 B
▼ ¥ ×) /// Variables
e comrandi here 🕐 (1994,179 Gubervalions Leau)
ommad jr . infix uning "G'NEG_Dictionary.dot", uning "G'NEGA_STOS_UTTARAURAD.dat") clear Name Liber
Copercised and a constraint of the constraint of
ee same to 2,410 to 2,510 to 2
e "Glfample data_setdstr. district 3-4 town_vilage
NY MANO are rebuil 5-7 ward oddio
Internet and a set from the set of the set o
Animonia Di Collaria della mandana della d
الله المراقب ال مقادم المراقب ا مراقب مراقب المراقب الم
HT Emox 1142 #-1-1 memory whity grow HT Emox 1142# F-11 memory whoty grow HT Market Emox 2000 memory whoty grow HT Market Emox 2000 memory whoty grow HT memory memory whoty grow HT memory memory<
Na, Journel 11 trans, village 1-11 maximum Milling group 11 secondaria emmedia. Milling group 11 secondaria maximum Milling group 11 secondaria maximum Milling group 11 secondaria maximum Milling group 11 secondaria Milling group
million 11 tota rillion million million million 11 tota rillion million million addoca kini graup 11 tota million million million addoca kini graup 11 tota million million million addoca kini graup 11 tota million million million ating mon 11 tota million million million ating mon 11 tota structure moletin moletin ating mon 11 tota structure moletin moletin ating mon 11 tota structure moletin moletin ating mon 11 tota structure 27-21 moletin
Na, mon 15 E Emer, VILSep 1-13 meximum Media A Ling group 15 E Meridia Ling Sing Sing Sing Sing Sing Sing Sing S
with your 11 to the yillars 1-11 menution. with your 11 wetch with your menution. with with with your 11 wetch with your build/cole with your 11 wetch with your With with your with with with your 11 wetch your With your with with with your 11 wetch your wetch your with with with your 11 wetch your wetch your with with with with your 11 wetch your wetch your with with with with with your 11 wetch your wetch your <
and your H1 took yillage 1-11 memory. deck throug you H1 witch yields 1-13 execution of top print, you H1 execution 1-13 execution of top print, you H1 execution 1-13 execution of top print, you H1 execution 1-13 model of top print, you H1 execution 1-13 model of top print, you H1 execution 1-14 model of top print, you H1 execution 1-14 model of top print, you H1 execution 1-14 model of top print, you H1 holds 1-14 Microbio of domain (0003) H1 holds 1-14 Microbio of domain (0003) H1 holds 1-14 model of domain (0003) H1 holds 1-14 model of domain (0003) H1 holds 1-14 model of domain (0003) H1 execution 1-13 model of domain (0003) H1 execution 1-14 model of domain (0003) H1 execution 1-13 model of domain (0003) H1 execution 1-13 model of domain (0003)
Number String Status String Status </td
Na, Mono FT <
Minggam HT 1000, VII.300, II.01 Minggam Mingam Mingam Minggam </td
Na Onio 11 11 11 11 11 11 11 11 11 11 11 11 11
Na Jook 1 ft for verside de 1-13 ster verside de 1-14 ster versi
All of the set state of the set set of the set set of the set of t
Na Onco III Constitute 1-13 Standard Constitute 1-13 Standard Environment (Standard Constitute 1-13 Standard Environment (Standard Constitute 1-13 Standard Constitute 1-1
A constrained with the second
N. Once Antigraphi (ft Elevis y 11/28 - 1-1) the web of the set
Subsymp 11 to town yilling 1-11 Interaction Interaction Sector Long your 11 to town yilling 1-13 Interaction Interaction Sector Long your 11 to encode 0.000 Interaction Interaction Interaction Sector Long your 11 to encode 0.000 Interaction Interaction Interaction Sector Long your 11 to encode 0.000 Interaction Interaction Interaction Sector Long your 11 to encode 0.000 Interaction Interaction Interaction Sector Long your 11 to encode 0.000 Interaction Interaction Interaction Sector Long your 11 to encode 0.000 Interaction Interaction Interaction Sector Long your 11 to encode 0.000 Interaction Interaction Interaction Sector Long your 11 to encode 0.000 Interaction Interaction Interaction Sector Long your 11 to encode 0.000 Interaction Interaction Interaction Sector Long your 11 to encode 0.000 Interaction Interaction Interaction
An own of the tory stillses 1-13 ended a long sing you dot a long you dot you



This is the one. We have already open one dataset. Either you clear this first, otherwise you simply go there and we have option. File, then import, text data in fixed format, there are two way giving, one is for our dictionary file, use dictionary file, click that. So, dictionary file as I told you already you have already saved as dot dct, just double click on it. It will save. Specification is not required at this moment because we are going through the dct file. Dct we have already given the specification. But if have not created a dct file then you have to enter the specification. We are going to guide you.

What is the dataset you wanted to extract, since our dct is for the Uttarakhand, specifically for the state, we have accordingly guide the particular fixed format data that is available for Uttarakhand? I have already shown you by state wise differences. We have only taken Uttarakhand here. what I will do, I will show you the Uttarakhand one. So, it is extracted one. We extracted it and we will show you here. This is the one Uttarakhand, we have opened it, isn't it? So, this is dot dat format.

If you submit, then, since a database was already been opened, it will be creating overlapping and some confusion, error will come, so replace the data in memory, whatever is there it will replace it and simply okay, it will extract. Your data is extracted with the help a dct file we created. This is your data and better to extract in string that then later on you can destring it and use it for your purpose. (Refer Slide Time: 24:32)

Main if/in
• Use dictionary file:
Browse
○ Specifications:
Fermular
examples.
2 lines 1: id 1-6 str name 7-36 2: age 1-2 sex 4
Text dataset filename: (required if not specified in dictionary)
Browse
Replace data in memory

Then second aspect, second approach you just look at carefully. If it is not there, likewise, this is the first one we have already experimented it and shown you the extraction.

(Refer Slide Time: 24:36)



Cut 🔄 🖾 Layout -	Calibri (Body) - 20 - J	x x → → H + H + # # # 18 - 11 Test Direction -	811000 Q. 🕞	🛕 Shape Fill *	Ma Find			
Copy Reset		Align Text -	ALLOGG - Arrange Quick	Shape Outline	Replace ·			
Side - Diete	Int	Paramanh 9	Pravino	Shape Energy -	e seed *			
Outline X		and the state of t					- 0	x
	Computer	Removable Disk (G)				• • Search R	amountile Dask (G1	0
AND TANK		Cody address					10 P	ź
parter i fersourte	Organize • • • Open	Burn "New folder					(III - III	•
	🚖 Favorites	Name	Date modified	Type	Size			- 1
	Desktop	Reports	7/13/2020 10:30 PM	File folder				
n an an a parte sen jere a parte fer fe	Downloads Recent Places	Stata15	6/18/2020 5:58 PM	File folder				- 1
100 - C	Ap Incomi Places	State_wise_Data_IN_ASCI_FORMAT	7/13/2020 10:30 PM 7/13/2020 10:31 PM	File folder				- 1
	📜 Libraries	CGA_STOS_UTTARAKHAND	6/8/2016 5-32 AM	DAT File	22.327 KB			
	Bocuments	EC6_Dictionary.dct	8/17/2020 12:31 A	DCT File	1 KB			
	🕹 Music	my_da.do	8/4/2020 2-25 PM	DO File	2 KB			
federal of Galaxy to Fast here the	Notices	EC6_uttarakhand_extracted.dta	8/17/2020 12:34 A	DTA File	21,573 KB			
Contro, Tar (ad) 7 you from realise of the city of on a fight your doaler raises when stage of a well court gathers a movie your detect for		exempte data_secuta	8/10/2020 5:18 AM 8/10/2020 5:01 AM	DTA File	48.152 KR			
erten vor gestellt in til dörang få in dett ges internor i samtiliter	🐴 Computer	practice_dataset_NSS73_block2.dta	8/6/2020 2:26 PM	DTA File	19,009 KB			
	💩 Local Disk (C.)	S NSS1-civ	8/17/2020 2:32 AM	Microsoft Office E	2.388 KB			
14 and 14	Local Disk (D)	Si nos	8/17/2020 2:40 AM	Microsoft Office E	3,351 KB			
- 1	Removable Disk (C)	Lecture 20- variable and value labels Lecture 31- Extraction of unit total in total	8/21/2020 1:58 PM	Microsoft Office P	1,587 KB			
La Productione and the second	The constraint more (0)	NSS1-tab.tab	8/17/2020 2:342 PM 8/17/2020 2:34 AM	TAB File	2.388 KR			
	🚯 Network		WATER CALIFORN		1,000 mB			
errere cities dil Alberty di fute cite adde control had australità andre della attacta aggi								
to de las la dela tel la del tel lan tel manage de encretado de la contra tel de la contra della tel manage de la contra de la contra della contra della								
	EC6A_ST05_UT	TARAXHAND Date modified: 6/8/2016 5:32 AM Date of	reated: 8/17/2020 3:24 PM					
	DAT File	Size: 21.8 MB						- 1
22 "Office Theme" 🧶					_	_	_	
ISI Cata Graphics Statistics					Bixo		- 2041	
ISI I Data Graphics Statistics		₽ 0 0 0 	6 7 8 9	0	Bksp		- 2041	0 h
551 1 Data Graphics Statistics □ □ · :::::::::::::::::::::::::::::::::	Une Esc - Tab	n o · ¹ 1 ¹⁰ 2 ¹⁰ 3 ¹⁵ 4 ¹⁶ 5 ¹ q w e r t y	6 ^{&} 7 [*] 8 ^{(†} 9) u i o p		Bksp		 I O I I Variables 	0 N
ISI I Das Gapher Satisfer IIII - Li - X - X - X IIII - Li - X - X - III IIII - III - IIII - IIII IIIII - IIII - IIII - IIII IIIII - IIII - IIII - IIII IIIII - IIII - IIII - IIIII IIIII - IIII - IIII - IIIII IIIII - IIIII - IIIII IIIII - IIIII IIIIIIII IIIIIIIIII	User Esc infit fix c Caps	■ 2 ► 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5	6 ^{&} 7 [*] 8 ⁽ 9 ⁾ u i o p		Bksp		 I O I I Variables Titer constitution 	
Image: State	User Esc infi fix c Shift		6 ⁸ 7 *8 ⁽ 9) u i o p j k i ⊧		Bksp b Shift		 Variables Titer variable Name 	s here
Image: State of the state o	Uer Esc - infli Caps Shift - cap - trab		6 ^{&} 7 [*] 8 ^{(*} 9 [†]) U i o p J k I j n m c p		Bksp Dc Shift		 Image: Second sec	s here
Image: Data Graphics Statistics Image: Data Graphics Statistics Image: Data Graphics Statistics Image: Data Graphics Y = X × 1 Image: Dat	Uer Ex - Tab Tab Shift Caps Shift Cirl 4	. • · 1 • 2 • 3 5 4 • 5 ^ 9 • • 1 • 7 • 7 • 7 • 7 • 7 • 7 • 7 • 7 •	6 ^(k) 7 [*] 8 ^{(L} 9 ⁽) u i 0 p j k n m < > . At = Cut		Bitsp I V Dr Shift + Fri		 ▲ Solution Variables Mane State district 	s here
Complex Subsets C	Uver Esc Tab Infin Caps Shift caps Shift caps Chift cap Chift caps Chift caps Chift caps Chift caps Chift caps Chift c	Image: Contract of the second secon	6 * 7 * 8 * 9 10 i o p 1 k l ; n m < > At m Cti		Bitsp A Do Shift + Fri		 Variables Filter variable Name state district tehsil 	s here
351 I Das Gapta Stating II Das Gapta Stating	Uer Esc - Tab Infi c Gaps Shift Esc - F at C - C		6 ⁶ 7 8 ⁽ 9 ⁾ u i o p j k i n m · · · · At ⁸ Cui	0	Bitsp A Do Shift - Fr		Variables Variab	s here
151 Data Graphica Statistica Tota Coupling Statistica Tota Coupling Statistica Tota X Tota X Tota X Inspaces, distance, 1 Statistica Statis Statis <td>Ure State and State Shift Caps Shift Ca</td> <td></td> <td>6 ⁰ 7 * 8 [[] 9 []] U I 0 P J k I <u>F</u> In Im (* F . Jat Iⁿ Ori</td> <td></td> <td>Bitsp Bitsp Shift P Fri</td> <td></td> <td>Variables Tater vanable Name State Sehoil</td> <td>s here Label</td>	Ure State and State Shift Caps Shift Ca		6 ⁰ 7 * 8 [[] 9 []] U I 0 P J k I <u>F</u> In Im (* F . Jat I ⁿ Ori		Bitsp Bitsp Shift P Fri		Variables Tater vanable Name State Sehoil	s here Label
Date Graphics Statistics Date Graphics Statistics Date Graphics Statistics T + 1 + 1 Date Graphics Statistics T + 1 + 1 Date Graphics T + 1 + 1 Date Graphics T + 1 Date Graphics D	Unr Sc Control	Image: Control of the system Image: Control of the system <td< td=""><td>6 ⁶ 7 <mark>6 8 ⁶ 9 1</mark> 10 1 0 p 11 k 1 1 10 m <u>6 p</u> 10 m <u>6 p</u></td><td></td><td>= Bisp A De Shift * Fri</td><td></td><td>Variables Fifter vanishe Name Sate district tehoil</td><td>s here</td></td<>	6 ⁶ 7 <mark>6 8 ⁶ 9 1</mark> 10 1 0 p 11 k 1 1 10 m <u>6 p</u> 10 m <u>6 p</u>		= Bisp A De Shift * Fri		Variables Fifter vanishe Name Sate district tehoil	s here
Constant of a second seco	Une Solution Control of Control o	Image: Constraint of the second sec	6 ⁽⁰ 7 [°] 8 ⁽¹ 9 ⁽¹) u ii 0 p j k U ∏ n [m (<) ∧t m (cri		Biksp A Do Shift * Fn		Variables Variables Triter sanst/e datect setei	s here
31 1 1016 Graphics Statistics 11 2 1 2 11 2 1 2 1 11 2 1 2 1 2 11 2 1 2 1 1 1 11 3 1 </td <td>Uner Soc</td> <td>Image: Constraint of the second sec</td> <td>6 ⁰ 7 2 8 9 9 10 1 0 19 17 14 17 17 14 17 14 18 19 14 18 19</td> <td></td> <td>Biksp I \ Dc Shift * Fri</td> <td></td> <td>Variables Variables Triter sounder state district tehoil</td> <td>s here</td>	Uner Soc	Image: Constraint of the second sec	6 ⁰ 7 2 8 9 9 10 1 0 19 17 14 17 17 14 17 14 18 19 14 18 19		Biksp I \ Dc Shift * Fri		Variables Variables Triter sounder state district tehoil	s here
Date Graphical Statistics Date Graphical Statistics Date Graphical Statistics Date Graphical Statistics T + 2 + 2 T + 2 + 2 T +	Constraints of the second	Image: Constraint of the second se	6 ⁰ 7 ¹ 8 ¹ 9 ¹ U 1 0 p J k 1 <u>1</u> n m 7 1 Ar a ori	0 - 	Bisp Shift • Fri		Variobles Variobles Trate vanable Rame state setoil	s here
31 1 10.0 Copyrect: Statement 10.0	Image: Section 1 Image: Section 1	Image: Constraint of the second se	6 10 7 8 10 9 10 11 10 10 13 14 10 1 10 1	0	Bitsp A De Shift * Fri		Variables Trifer vander Name State dation tehoil	s here
ST C Did Graphic Statistic C Did Graphic St	Comparison of the second	Image: Control of the second secon	6 10 7 1 8 1 9 1 10 1 10 19 11 10 19 10 10 19 10 10 19 10 10 19 10 10 19 10 10 19 10 10 10 19 10 1	0 (()))))))))))))	Bitsp A Do Shift * Fr		 Variables The vanishe have the vanishe district sehoil 	s here
Compared Statement Compared Co	Constraints of the second	Image: Second	6 ^{(h} 7 ⁽ 8 ^{(l} 9 ^{(l}) u i o p j k l i n m r <u>r</u> At n ori	0 · · (1 · ·) - / · · - / · ·	Shift • Fri		Variables Variables Transme State district sehell	s here Label
Control Contro Control Control Control Control Control Control Control Control Co	Comparison of the second		6 / 7 / 8 / 9 / U i o p jj k / 1 - j k / 1 - j k / 1 - At n (cr)		Shift * Fri		Variables Variables Intervander Name state state tehsi Properties @ - + Variable.	s here
St Duta Graphics Statistics Duta Graphics Statistics Duta Graphics Statistics TV + TV	Compared and a second and	Image: Control of the second secon	6 ⁶ 7 <mark>6 8 ⁶ 9 ⁶</mark> 9 1 6 9 1 k 1 7 0 m 6 6 7 0 m 7 0 1 0		Bitsp I L Du Shift * Fr		Variables Variables The variable State state state state state state state state variables Name	s here Label
St Image: Souther Control Image: Souther Control Image: Souther Control<	Construction C	Image: Constraint of the second se	6 ^{(k} 7 ⁽ 8 ^{(l} 9 ^{(l}) 0 ^{(l} 0 ^{(p}) 1 ^{(k} 1 ^{(l}) 0 ^{(m} ^{(c}) 0 ^{(l}) 0 ^{(m} ^(c) 0 ^{(l})		Bitsp File Shift File File File File File File File File		Variables Variables Tole vanide Rane Sate Sate Sate Variables Variables Name Labei	s here Label
Common Construction C	the second	Image: Constraint of the second se	6 ⁰ 7 ¹ 8 ⁰ 9 ⁰ 2 1 0 p 1 k 1 <u>p</u> 1 m c 1 1 1 1 1 1 1 1		Suit		Variables Variables Inter-canade Inter-can	s here Label
St Data Graphics Statistics Data Graphics Statistics Data Graphics Statistics T + 4 T +	Construction of the second secon		6 ⁰ 7 ¹ 8 ⁰ 9 ¹ U 1 0 p J k 1 <u>1</u> n m <u>r 1</u> At n at		shit • friques		Variables Variables Variables Figure sandle Rame Setuil Properties Properties Name Care Setuil Variables Name Care Setuil Variables Care Setuil Variables Care Setuil Care Setuil Care Setuil Care Setuil Care Setuil Care Setuil Care Setuil Care Setuil Care Setuil Care Setuil Care Setuil Care Setuil Care Setu	s here Label
Construction C	Comparing the second seco	Image: Control of the second secon	6 ⁰ 7 ¹ 8 ¹ 9 ¹ u i o p j k i o p		Shit		Variables Variables Variables Variables Variables Variables Variables Variables Name Later Spen Spen Spen Spen Spen Spen Spen Spen	shere Label
Control Control Control Control Control Co	Comparison of the second				Filep		Vanables Vanables Vanables Vanables Properties Properties Name Later Setai Vanables Name Later Setai Vanables Name Later Name Later Name Later Name Later Name Later Name Name Later Name Name Name Name Name Name Name Name	s here Label
100 100 0 summer 0 summer 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 </td <td>Comparing a second second</td> <td>Image: Constraint of the second se</td> <td>6 7 8 9 1 0 0 0 1 k 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td> <td>o en la constanta da la consta</td> <td>Shift - Fr</td> <td></td> <td>Variadas Variadas Nore Nore Be Otto Be Variadas Be Be<!--</td--><td>s here Labet</td></td>	Comparing a second	Image: Constraint of the second se	6 7 8 9 1 0 0 0 1 k 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0	o en la constanta da la consta	Shift - Fr		Variadas Variadas Nore Nore Be Otto Be Variadas Be Be </td <td>s here Labet</td>	s here Labet
Control of the second sec	Comparing the second seco	Image: Constraint of the second sec	6 0 7 8 9 9 U 1 0 P J K T J At 1 CH At 1 CH C4_155_077200000 64, 0	0	shift a pri		Vanidas Vanida	s here Label
St C C C C C C C C C C C C C	Comparing the second seco	Image: Constraint of the second se	6 7 8 9 1 1 0 p 1 0 p	0	j Bicp		Vauldes Vauldes Interference Interference Interference Base Base <	s here Label
SI Dub courses Sources Tax Courses Sources Tax Courses Sources Tax Courses Sources Tax Courses Sources Tax Courses Sources Stanspre data parts Stanspre data parts Stans	Comparison of the second		6 0 7 8 9 0 1 0 0 1 k 1 - n m - At a ori		shit		Venables Venables <t< td=""><td>shee Label</td></t<>	shee Label
Control of the second sec	Comparison of the second	Image: Constraint of the second sec	6 0 7 8 9 9 9 1 0 9 1 k 1 1 0 0 0	a	Solution		Vasalas Vasala	Shee Label 13941 2695
Clamping Synthese	Compared and a second a secon	Image: Constraint of the second se	6 7 8 9 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0	o 			Values Values International statements International statements <td>s here Label 1 394.1 2.633 66M</td>	s here Label 1 394.1 2.633 66M

	state[1]							
	state 61	anie i	th#11				Variables	
1	0.5	01	006				A Discoutie	
2	05	01	006				Ch Harris	
3	05	01	006				Marne Ca	Lapei
4	05	02	003				State	
5	05	02	003				Control (Control)	
6	05	02	004				🗶 unal	
7	05	02	005					
8	05	03	001					
9	05	0.3	001					
10	05	04	005					
11	05	04	005					
12	05	04	005					
13	05	04	005					
38. 	05		005					
15	03		005					
17	03		445				Variables Snapsh	ots
18	03		005				Properties	
	05	04	007				 Variables 	
		05	003				Name	state
20			003				Label	
20	05	05					Type	10/2
20 21 22	05	05 05	003					
20 21 22 23	05 05 05	05 05	003					
20 21 22 23 24	05 05 05 05	05 05 05	003 003 004				Value label	
20 21 22 23 24 25	05 05 05 05 05	05 05 05 05	003 003 004 004				Value label Notes	
20 21 22 23 24 25 26	05 05 05 05 05 05 05	05 05 05 05 05	003 003 004 004 004				Value label Notes	
20 21 22 23 24 25 24 25 24 25 26 27	03 05 05 05 05 05 05 05	05 05 05 05 05	003 003 004 004 004 004				Value label Notes • Data	
20 21 22 23 24 25 26 27 28	03 05 05 05 05 05 05 05	05 05 05 05 05 05 05	003 009 004 004 004 004 004				Value label Notes • Data Filename Label	
20 21 22 23 24 25 26 27 28 29	03 05 05 05 05 05 05 05 05	05 05 05 05 05 05 05 05	003 003 004 004 004 004 004 004				Value label Notes • Data • Filename Label Notes	
20 21 22 23 24 25 26 27 28 29 30	05 05 05 05 05 05 05 05 05	05 05 05 05 05 05 05 05 05	003 003 004 004 004 004 004 004				Value label Notes]
20 21 22 23 24 25 26 27 28 29 30 31	05 05 05 05 05 05 05 05 05 05 05 05	05 05 05 05 05 05 05 05 05 05 05	003 003 004 004 004 004 004 004				Value label Notes • Data • Fiename Label Notes Vanables Observations • Gue	3 394,179 3.63M

The second approach is, use the infix command. Infix command is useful for fixed format data and generally the database is available in fixed format nowadays. So, the sample the way how the data looks like you can also understand yes this is in fixed format data. So, that way you can also observe. So, use infix command without data dictionary. If it is not there, then data dictionary you have not created you have the byte space of the particular variable, their position and you can also extract it.

Suppose, and generally this is one, this approach is done when you have a limited variable. You want limited variable extraction, but you can manually do it, no need to create a dictionary file. But it is always suggested that you please take the advance approach, dictionary is always suggested for you. But let us copy this format right now, simply copy. Copy infix from infix till clear. I am just copying it, starting till end. This is there. So, now it will work.

So, what I will do, I will change this here. So the next slide we have, still it is coming. Let me change to then visible automatic, if I do it, it will come, now copy, you will need to end this, you can able to copy from here. If you go on that particular page, you will be able to copy that command. We are deliberately doing it to save your time. And the crux is that we know the position. If you do not know, please do not simply copy and run, it will not work.

Now we have copied it and we will operate this in our Stata, here is our Stata. What I will do, I will clear the data first. First clear, already clear is there in the command. How many variables

we have copied? We have copied state 1, then 2, then tehsil 3. Another variable also EC6A underscore ST05. But this is the file name of that particular.

So, we have selected the variable first, whichever the variable you want to select, you want to extract and using is very important, using the source of that particular file. Source is important. You cannot just simply assume that this is your right source, because we are operating from another system. We have given a standard command as per our own laptop. I have to change this using file. Using source has to be given.

So, I am deliberately deleting it for your better understanding. Using file, where is our using file available, it is available in our data, here. So, using file, what is that using file, this is the one. This is the using file, Uttarakhand, not extracted, but this is the original data, this is the one, yes. The raw data, we want the raw data to be extracted. What is the file name? You need to copy this, right click here, copy address then I am going to paste it. So, right click here and paste. So, the path name has already copied. What next comes, you require the file name. So, file name, you just simply copy it, right click and copy. It has already been copied. So I am going to paste it here.

What is this file, it is in dta, dat file? So original, so file, how to know that dat is there already. It is clearly written dat file. So, you have to mention that it is in dat file. if you enter your data will be ready. So, your data is ready like this. it has been extracted. mark it very carefully. we have extracted based on our 3 variables. So, three variables information is there for you in our window, only 3 variables. And this is interestingly, you have not created a dictionary file.

(Refer Slide Time: 30:07)



I am going to guide you that what other options are there in front of us. We have only the limited operations are there for your clarity. We have already guided you this. Stata reads string variable state from column one through column two that I have already told you please mark my directions very carefully. Since the infix recognizes a data item by its column range not by a delimiter, as I told you already from the beginning, there is no delimiter. Column range is mentioned in our layout file. Layout file we already guided.

Users may skip variables or ignore the order of variables. So, order of variable need not be mentioned because we are specifying their order by the position one through two or through 5 or 6 whatever is given in the layout file accordingly. So, order is not necessary. It is very important to note. You may also select observations to be read using the in or if qualifier there. If those qualifiers are mentioned, then we will also mention accordingly.

So, like in our data you want only 100 observations, so this is basically if qualifiers command we are going to give it. So, infix then the same format, everything is same except we are adding here one thing that is this. I am going to select this for your use that 100, 1 through 100. 100 observations it will automatically extract to 100. So I am not going to experiment right now. You can do it on your own. If you have difficulties, please raise.

(Refer Slide Time: 31:50)



The third method is also important. Use infix command with data dictionary. If there are many variables in a complicated format, you can benefit from writing a data dictionary. This approach is highly recommended for the purpose of data management.

(Refer Slide Time: 32:08)





ļ	2.0.		Drawing Topis Lecture 21-1	Extraction of unit level data in stata - Microsoft Rower	loint	- • X
	Home Insert Design Anim	ations Side Show Review	View MathType Format			U.
	▲ Cut □ Layon Parte ✔ Format Painter Clipboard Slide - Slides	Calibri (Body) - 24 - A	A (♥) □ + □ + □ + ● ● □ + → □ Align Ter A + ● ● ● ● ■ ■ ■ ■ + → ⊕ Convert 1 A + ● ● ■ ■ ■ ■ ■ + ⊕ Convert 1	etien * E \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \	Dape Fill + B Find Dape Outine + S Find Dape Outine + S Find Shape (Herds + S Select + D Editing	
	Slides Outline X					
		Organize • Open	Removabl Dum Copy address Dum Edit address Name Delete history	Date modified	Type [*] Size	Search Removable Datt (G)
		Desktop	🗼 Reports	7/13/2020 10:30 PM	File folder	
	10 Discontinue au ris promo activación activación Discontiniza des antides contra de acceleration	Downloads Recent Places	Stata15 State_wise_Data_IN_ASCIL_FORMAT Supplementary_OOCUMENTS	6/18/2020 5:58 PM 7/13/2020 10:30 PM 7/13/2020 10:31 PM	File folder File folder File folder	
	the second secon	💢 Libraries	EC6A_ST05_UTTARAKHAND	6/8/2016 5:32 AM	DAT File 22.327 KB	
	the body the star	Bocuments	EC6_Dictionary.dct	8/17/2020 12:31 A.	DCT File 1 KB	
		💰 Music	my_do.do	8/4/2020 2:25 PM	DO File 2 KB	
	Directory of second sectors in contrasts of	S Pictures	EC6_uttarakhand_extracted.dta	8/17/2020 12:34 A	DTA File 21,573 KB	
	 b. a definition and the gap of setting of the paper in order of or taking b. the gap and computing to the state of the taking 	S viceos	Example data_set.dta	8/6/2020 3:18 AM	DTA File 17,754 KB	
	weiter Faanfann - N - Gale 10 - Boler 14 - Biel 17 - Ang Versjellig, Fellewerd is 148 - 16	Computer	practice_dataset_IHDS.dta	8/10/2020 5/01 AM	DTA File 48,152 KB	
	exclusion and an all memory	Local Disk (C)	Sal NSS1.cm	8/0/2020 2:20 PM 8/17/2020 2:23 AM	Microsoft Office E 2188 KB	
		 Local Disk (D.) 	50 ms.	8/17/2020 2:40 AM	Microsoft Office E. 3351 KB	
	2 Tel Miles		Lecture 20- variable and value labels	8/21/2020 1:58 PM	Microsoft Office P., 1,587 KB	
	um info germane una ambéricame. Bir funcion travalente la constante funciona acconstante. Nen una constantentes.	Removable Disk (G)	Secture 21- Extraction of unit level data in st	ata 8/21/2020 3:42 PM	Microsoft Office P., 1,957 KB	
		🕵 Network		8/17/2020 2.34 AM	TAB file 2.388 KB	
	20 Stde 30 of 22 "Office Theme" 3	EC6_Dictionary DCT File	rdet Date modified: 8/17/2020 12:31 AM Da Size: 565 bytes	te created: 8/17/2020 3:24 PM		0 8
		e 💙 🔜 🕨	<u> </u>			8/21/2020



So, like in the infix command itself you can add the dictionary, the complete extraction command. So, infix using dictionary file, two times using dictionary because you want dictionary file. This is the command we can give it in our Stata. You can copy this, except this three slash, this is not required. Then you simply copy and paste it. But make sure that you are using the right using file as well as write dictionary path. This is very very important. But if you do not use this correctly, it will be erroneous. We can experiment that as well, because infix you can copy that. You can copy this. Infix, we are copying it.

Most importantly the parenthesis is very very important. Raw data files name should be always in parenthesis. That is another important instruction. Otherwise, Stata will throw an error. We have now copied it. So, what I will do, we will go to Stata. So we will paste it here. This can be avoided. We need to change the path. So, path we will get it the way we did. So the path we did like we have to copy the path of that from the data file, first the dct file should be understood. First one is dct. Here is the dct. So dct, you copy that. We are copying the path. So, we are pasting it in our Stata. So, we are putting the path here.

We are getting the dct file to be copied. Complete dct file to be copied. We have copied also this and we have put. Dot dct is already in the file name. So, we have to keep that plus we need to add dot dct, because path name has to be correctly retain. So, two times dot dct, do not get mislead by two times, because in our file name itself we have written as dot dct. we are using the using file that is the raw data file. Raw data file format I already told you.

The data it is here. It is the raw data. We are copying the path of that raw data once again, copying the address for you. And the address is in G file, G colon forward slash then the file name. So, we are putting the file name with extension, the dat extension. It is in dat file, comma then clear. Clear we have already given whatever is already saved in the memory, this will be deleting and giving us a final extraction.

You must start with an empty dataset is still saying that, I think, so the problem is coming. The clear, first of all we cleared the command. The dct, it will read. I think let me first read, no, I think there is one thing missing from the. This is the one. You need the dct file. you have seen this once again. Where is that, So, we are going to use the same command once again. Then parenthesis is clearly written. Parenthesis will be read for sure.

What happens since it is, what we do we will first discuss, no, we will first close the Stata and we will reopen it. Because you have already stored the extraction, we will reopen the Stata and do it. It is possible. We already did it on our own. So, this is the one. We are going to open it. I think if you paste it, infix using then the dct path name, dct path name is here, the G, it is in G, then dct name should be also copied.

This is the one. Entire has to be copied right now and we have copied right now and pasting it here with dot dct. Then it ends with inverted comma then using the file name, the file name within bracket is important and copied the path again and file name you can. This is the one, dot dat bracket close. The file this dictionary not found.

So, two times dot it does not take. So, we have finally experimented it with the right command. let me correctly specify you and mention you and guide you how you have done it. You have extracted it. So, the file name should be correctly mentioned. So, the .dct should be written, but in the file name two times dct is creating some problem. So, better to mention, file name should not be entering with a special character. So, it has to be file then ended with your extension. Extension two times you are entering, so there was some problem. We did it and you can follow accordingly.

(Refer Slide Time: 40:08)



So, another approach is simply, this is the one if you enter it you extract it. Otherwise, in the dictionary itself if you enter in fixed dictionary using this file name here itself and after mentioning all those dictionary commands, dictionary syntax if you run, it will also execute with the extraction. So, these are all converting to the result and giving you the right data.

So, what I wanted to mention at the end for this extraction is that infix command is very very important and all the dataset available in infix command. There are approaches you can follow. We will always suggest that you use the dictionary, dot dct approach and that is going to be very

useful and you have to make it very confidently and to be very handy with the dictionary and that will be very, very useful for your better result and quick results.

And this is all for today for extraction. In the next class we will have important aspects as well called combining datasets, merging and appending that is very very important for researchers in social science and other discipline as well. With this, let me close here friends. Thank you so much.